Alphabetical Order, the Dewey Decimal System, and Google

David W. Lewis
Dean of the IUPUI University Library

Read on Tuesday February 21, 2006 at the regular meeting of the Indianapolis Literary Club, Park Tudor School, Indianapolis, Indiana

Introduction: Three Revolutions

Revolution One: The Invention of Printing

In Europe just after the turn of the fourteenth century a series of technological developments coalesced to create the innovation that arguably made possible the modern world and the modern mind – the invention of printing.

We know that paper and wood block printing developed in China — paper as early as 100 B.C. and woodblock printing as early as the sixth century. It is unclear the exact route by which these innovations came to Europe, but by the early fifteenth century moveable wooden type was being used to print on paper.

There has been some dispute about who actually invented the printing press, but Johannes Gutenberg, who built his press around 1436 and in the 1450s produced the first printed Latin Bible, is inevitably given the credit. Gutenberg's contribution was the combination of a press used at the time for binding codices and metal dies and matrices that created re-arrangeable type. He also was an early user of oil-based inks.

With printing the world changed and changed radically. The scribal culture, which had held sway since the invention of writing between 5,000 and 6,000 years ago, was washed away almost without a trace. The church and the universities no longer had exclusive control of the means of knowledge production. Arguably printing played a key role in the Renaissance, the Protestant Reformation, and the Copernican Revolution in science. Others have argued that it weakened the Papacy and lead to the fall of the Byzantine Empire to the Turks and that indirectly it lead to the European discovery of the Americas. What is clear is that knowledge escaped the grasp of the elites and became more widely available. It is difficult to assess the expansion of knowledge that resulted from the

invention of printing, but as one indication, we can look at WorldCat, the largest database of books, which contains the combined holdings of thousands of the world's largest libraries. It contains 102 records that include the word "Bible" for books that were produced between 1400 and 1450. For works produced between 1450 and 1500 there are 1,120 records — a better than ten fold increase. Between 1500 and 1550 there are 5,800 such records — a 50-fold increase from the same period a century before.[1] We should also remember that the typical press run of an edition between 1450 and 1500, as documented by Elizabeth Eisenstein, was 200 to 1,000. An "edition" before printing, if the word can even be used, was one.[2] This comparison is difficult because this attempt to measure manuscripts production, like most others, is subject to many methodological problems, but it certainly shows the trend.

What is beyond doubt is that printing increased the production of books and other written material by several orders of magnitude. Many more people had access to much more knowledge and the world could never be the same again.

Revolution Two: The Industrialization of Printing

In the nineteenth century there was a similar, though less appreciated, revolution in how knowledge is conveyed. Friedrich Koenig and Andreas Friedrich Bauer invented the steam-powered printing press in 1812 and it was first used — to print an edition of *The Times* of London — on November 29,1814. At about the same time steam-driven paper making machines, which could make paper with fibers from wood pulp, were developed. Printing was industrialized.

The result was a major transformation of the nineteenth century society. Before this era books and newspapers were luxury objects and illiteracy was the norm for the majority. Through the nineteenth century, because of the introduction of cheap paper and the steam-powered printing press, schoolbooks, dime novels, and newspapers became widely available to most people in industrial societies. Cheap wood based paper also meant that keeping personal diaries or writing letters ceased to be reserved to a privileged few. Also at this time, the fountain pen and the mass produced pencil became widely available.

With the wider availability of books and the increase in literacy, public libraries, many funded by Andrew Carnegie, became a required cultural institution in American communities. With the Morrill Act of 1862 Land Grant Universities were established and university libraries grew as industrial printing made possible huge increases in the published scholarly and scientific literature.

Again, knowledge became more widely available, and was democratized as it became available to more people. Again the size of this increase is difficult to gauge precisely, but if we look at the comprehensive bibliographies of American publishing compiled by Charles Evans and Richard Shoemaker, there were 455 books published in the American

[1] WorldCat searches were conducted on February 6, 2006.

[2] Eisenstein, Elizabeth L., *The Printing Revolution in Early Modern Europe*, Cambridge: Cambridge University Press, 1983, page 9.

colonies in 1773. In 1823 there were 4,490, and by 1846 there were 7,783 — seventeen times more than 73 years before.[3]

One result of these developments was that by the end of the nineteenth century, as it is often joked, everyone in Indiana had become an author. Another result, one might argue, was the foundation of this club.

Revolution Three: The Internet

In our time, a third revolution in the way in which knowledge is produced and distributed is underway. The Internet and the World Wide Web, even in their infancy, have already produced profound changes, and we can reasonably expect social change as profound as that wrought by the previous two revolutions to result.

Vint Cerf and others working with the Defense Advanced Research Projects Agency, or DARPA, developed the TCP/IP protocols in the late 1970s and they became the standard for all network traffic on the ARPANET on January 1, 1983. By the mid 1990s the network was opened to commercial providers. At about the same time Tim Berners-Lee combined the idea of hypertext with the Internet and created the protocols and programming that made possible the World Wide Web. Version 1.0 of Mosaic, the first widely used graphical web browser, was released on April 22, 1993. In November 1997, a study estimated that the Web contained 200 million pages.[4] In January 2005 it was estimated that there were 11.5 billion pages — a nearly 60-fold increase in a little over seven years.[5]

Importantly, the Web allows individuals with only minimal expertise to create pages and to "publish" what they wish. A study of how much information was in existence in 2003 estimated that between 327 and 1,634 terabytes of printed material existed and that between 80% and 85% of this was in the form of office documents. This leaves between 49 and 326 terabytes of information in other printed formats — books, newspaper, journals, etc. The same study estimated that the Web in 2003 contained 532,897 terabytes of data of which about 82% was e-mail. This leaves 95,920 terabytes of data as content on the Web in 2003.[6] If this is approximately correct, the Web in 2003 contained between 300 and 1,840 times as much information as all world's printed material. As a point of reference, the Library of Congress, the largest library in the world, contains

---

[3] See Evans, Charles, American Bibliography, New York: Peter Smith, 1941, A Checklist of American Imprints for 1823, compiled by Richard H. Shoemaker, Metuchen, NJ: Scarecrow Press, 1968, and A Checklist of American Imprints 1846, compiled by Corl Rinderknecht and Scott Bruntjen, Lanham, MD, 1997.

[4] Bharat, K and AZ Broder, "A Technique for Measuring the Relative Size and Overlap of the Public Web Search Engines – WWW7 / Computer Networks, 1998. http://net.pku.edu.cn/~wbia/2004/public_html/Readings/web%20graph/Estimating%20the%20Relative%20Size%20and%20Overlap%20of%20Public%20Web%20Search%20Engines.pdf (February 4, 2006).

[5] Gulli, Antonio and Allessio Signonni, "The Indexable Web is more than 11.5 billion pages," http://www.cs.uiowa.edu/~asignori/web-size/ (February 4, 2006).

[6] "How Much Information? 2003," http://www.sims.berkeley.edu:8000/research/projects/how-much-info-2003/execsum.htm

about 10 terabytes of data. The Web in 2003, thus, contained the equivalent of 9,500 Libraries of Congress.

Once again knowledge had broken free from social and technical constraints. It had dramatically increased in quantity and become more democratic, easier to produce and easily accessible word-wide.

Finding It

There is much that could be said about these three revolutions, and tonight most of it will go unsaid.

The task I have set for myself is to consider how the quantum leaps in knowledge that these three revolutions engendered forced the development of new tools to organize it. As Jerry Campbell, the Dean of the University Library at the University of Southern California, has said, it is a, "fundamental truth of human knowledge: as the volume grows, principles of organization must be applied in order for knowledge to be used."[7]

What follows will be an examination of three technologies: alphabetical order, the Dewey Decimal System, and Google. We will consider each in turn and examine how each was a response to the expansion of knowledge brought on by the three revolutions I have just described.

Alphabetical Order

Between 5,000 and 6,000 years ago the first writing systems were developed in Sumer and Egypt. The oldest writings that have come down to the present day are inscriptions on clay tablets made by the Sumerians in about 3100 B.C. In Egypt the common reed Papyrus, parts of which were used for food and fuel, and which was also used for making boats, sails, mats, and cloth was adapted as a writing surface. These early writing systems began as pictorial systems, but with time the signs became stylized and lost their primary pictorial values and developed as ideograms, and syllabograms. The Phoenician alphabetic script of 22 letters was used at Byblos as early as the 15th century B.C. This method of writing, which was later adopted by the Greeks and is the ancestor of the modern Roman alphabet, revolutionized writing.

An alphabet is a complete standardized set of letters each of which roughly represents a phoneme of the spoken language. Alphabetic writing systems worked better than their predecessors because of their flexibility and economy. Alphabetic systems are far easier to learn and can be much more easily transferred from one language to another, in fact after its invention, the use of alphabetic systems spread so quickly that it is difficult to trace their history. Importantly for this discussion, the alphabet is an invariant series — A always comes before B, etc. This last attribute, as we shall see, was not fully applied

---

[7] Campbell, Jerry D., "Changing a Cultural Icon: The Academic Library as Virtual Destination," *EDUCAUSE Review* 41(1):20 January/February 2006. http://www.educause.edu/apps/er/erm06/erm0610.asp (February 4, 2006).

for some time, but this role as an ordering and finding device has become central to how we organize knowledge.

From the beginning alphabetic order was a teaching device for scribes, but it was used for little else. When writers, from the Greeks to medieval Christians and in literate cultures from China to Islam, wish to create compilations of knowledge they did so thematically. There were different thematic hierarchies from the Confucian classics to the seven liberal arts of the Greek academy to a variety of other schemes in Islam and Western Europe that attempted to combine religious and secular knowledge.

Before the invention of printing the only use of alphabetic order was in glossaries. In the early middle ages lists of difficult but important words began to appear as aids to the teaching of Latin. Over time the words in these compilations began to be arranged by their first letter. By the eighth century second or third level ordering was in use, but it took 400 more years before complete alphabetic order was common. But this use of the alphabet as an ordering system was at the margin. At a time when the intellectual challenge was to integrate, scholars of the time must have found the fragmentation and scattering that results from alphabetical order to have been distracting at best and an offensive at worst.

But as was the case with so much else, printing changed this as well. As Tom McArthur, whose book the *Worlds of Reference* was the inspiration for this essay, states:

> Although some properly alphabetic works appeared before Gutenberg printed his first book, the printing press seems to have been the factor that changed everything in favor of non-thematic ordering. Compositors were constantly re-shuffling the letters of the alphabet around as small hard metal objects in trays and in composites. They and their associates — which included many writers who were wont to frequent print shops — became as a consequence increasingly at home with the convenience that the alphabet offers as an invariant series. Where scholars and copyists had previously been unaccustomed even to thinking of words and parts of words alphabetically, printers were now spending a great part of their time doing nothing else.[8]

The widespread use of alphabetical order did not though appear overnight, but by 1600 concordances, indexes, and ABC wordbooks were common. Thus began the history of the alphabetically arranged reference books. The dictionaries of Samuel Johnson, published in 1755, and Noah Webster, published in 1828, and their successors, most notably the mammoth *New English Dictionary on Historical Principles*, or as it is more commonly known, the *Oxford English Dictionary* or even more simply, the *OED*, attempted to set forth the correct form of the language. In 1728 Ephraim Chambers produced the *Cyclopedia, or Universal Dictionary of Arts and Sciences*. This served as the model for the encyclopedias of Denis Diderot, published in 35 folio volumes between

[8] McArthur, Tom, *Worlds of Reference: Lexicography, Learning and Language from the Clay Tablet to the Computer*, Cambridge: Cambridge University Press, 1986, page 77.

1751 and 1780, and the *Encyclopedia Britannica*, first published as a three-volume work between 1768 and 1771 by a group of printers and scholars in Edinburgh who called themselves the "Society of Gentlemen in Scotland."

Printing created the capacity to rapidly produce texts in multiple copies. Before printing a scribe or scholar could spend months copying a text. Knowledge could be absorbed at a slower pace because there was less of it. After the invention of printing the amount of knowledge available increased and tools to assist in its management and use were required. Alphabetic order became an essential component of these tools.

McArthur describes the difference between thematic and alphabetic organization, "This dichotomy is far-reaching, however, because it operates first at a real practical level in terms of how works of reference are used and also at an ideal and theoretical level with regard to how information is best presented and understood."[9]

What is interesting to me is to consider not just how we shape these tools, but to consider how these tools shape us. Some communications theorists argue that alphabetic writing systems influenced the societies that used them and promoted the abstract skills of analysis, coding, and classification, and that these skills in turn lead to the development of such things as codified law, monotheism, abstract science, and deductive logic. It is difficult to prove any of this, but if this thinking has any merit, it seems equally likely that the use of alphabetical order as the central organizing mechanism of knowledge could have had similar cognitive impacts.

The Dewey Decimal System

In the second half of the nineteenth century, confronted with the rapidly expanding universe of printed materials made possible by the industrialization of printing, libraries faced a crisis. The systems they had used to organize and locate materials were failing. At the time most libraries arranged their collections in alcoves or rooms and each book was assigned a particular location on a shelf in an area related to the book's subject. As a result each library's system was unique. This approach worked when collections were small and grew slowly. Indexes, usually author and title lists were either hand written or printed lists so revisions were difficult. As collections grew more quickly and became larger both of these strategies were stretched to their limits. New approaches need to be devised.

While many individuals contributed to the development of the tools and techniques that made it possible to deal with the increases in knowledge brought on by the industrialization of print, the person who generally receives the most credit is Melvil Dewey.

Melville Louis Kossuth Dewey was born in the small upstate New York town of Adams Center, near Watertown, on December 10, 1851. He attended Amherst College and to help pay for his studies he worked in the library. Dewey remained at Amherst after his

[9] McArthur, Tom, *Worlds of Reference,* page 80.

graduation in 1874 and worked as a librarian. In 1876, at the age of twenty-five, he published *A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library,* in which he outlined what became known as the Dewey Decimal Classification system. That same year he moved to Boston and founded the Library Bureau a company founded as he said, "for the definite purpose of furnishing libraries with equipment and supplies of unvarying correctness and reliability,"[10] With R.R. Bowker and Frederick Leypoldt he founded the *Library Journal*, and helped to establish the American Library Association which held its first meeting that year.

Dewey was a driving force in American librarianship in the last quarter of the nineteenth century. In 1883 he became librarian of Columbia College in New York City, and there set up the School of Library Economy, the first institution for training librarians in the United States. The school was moved to Albany, as the State Library School in 1889 when Dewey moved to become director of the New York State Library. Dewey was also an advocate for spelling reform, a venture in which, sadly, was not particularly successful. He is responsible for the American spelling of "catalog" and managed to shorten his first name from "Melville" to "Melvil", but failed to gain acceptance for the change from "Dewey" to "Dui". Dewey was headstrong, a racist, anti-Semite, and he opposed woman's rights. These views contributed in no small part to his departure from both Columbia and the New York State Library and to business problems in his later life.

Dewey's classification system was, though, revolutionary, and his drive to reform and standardize library practice was critical to the establishment of the profession of librarianship in the United States. His work made possible the great print libraries of the twentieth century.

The Dewey Decimal Classification organizes the contents of a library based on the division of all knowledge into ten groups. Each group is assigned 100 numbers beginning with the general works in the 000s and concluding with history, biography, and geography in the 900s. The ten main groups are in turn subdivided again and again to provide more specific subject groups, for example the history of Europe is placed in the 940s, the history of England under 942, the history of the Stuart period at 942.06, and the history of the English Commonwealth at 942.063. This system solved the problem of fixed shelf location. When a Dewey number is combined with a Cutter for the author it has a fixed place in sequence, but any number of books can be inserted into the sequence. Dewey's classification system places books on similar topics in proximity to each other. The use of decimals was also critical to the success of the system for it could thus be expanded, at least in theory, indefinitely.

Practically, Dewey's classification system was only one part of the solution to organizing the growing library collections of the second half of the nineteenth century. A second, innovation was required to replace the book catalog and its inherent limitations. This innovation was the card catalog. Today card catalogs, while viewed by a few diehards nostalgically, seem quaint and outdated. But at the time of its introduction this was an inspired and essential technology. As you might expect, Dewey was involved. In 1877,

[10] From the Library Bureau website http://www.librarybureau.com/melvil.html (January 5, 2006)

largely at his urging, the American Library Association set the standard for card size — at 7.5 x 12.5 cm — which allowed for the production and sale of cards and cabinets. The Library Bureau was formed to do just that. The card catalog allows, again, at least in theory, for an infinitely expandable alphabetically arranged list of authors, titles, and subjects. An expandable standard classification scheme and a technology for building expandable catalogs lay the framework for library practice, which survived until well into the second half of the twentieth century.

Dewey's scheme is also interesting because it is based on the botanical metaphor of the tree as model of organization. This model was of course dominant in scientific classification and is based on the evolutionary assumptions of Darwin. It also reflects an attempt to impose a thematic hierarchy on all knowledge.

Interestingly, a similar effort to impose a thematic hierarchy on language lead to the creation of one of the most commonly used reference books — the thesaurus. What is today most often used as a means of enriching vocabulary or helping to solve crossword puzzles, began as a much more ambitious venture. Peter Mark Roget published his *Thesaurus of English Words and Phases* in 1852 and his aims were anything but pedestrian. He was attempting to develop a thematic hierarchy for the English language and to put all of its words in their proper place in this hierarchy. As McArthur puts it:

> His aim was to place a grid over reality, a kind of cartography of the mind, and label the appropriate nodes. He was aware, no doubt, that all such classifying is an exercise in putting continuums in containers, but went ahead with the work in the same spirit of optimism that animated natural history: because it was useful.[11]

It was only as an after thought that an index was added to the *Thesaurus*, and only when this happened did the work become successful. Roget's grand thematic hierarchy was made useful only because of alphabetical order.

Dewey's thematic hierarchy also gave way to practicality. The Library of Congress Classification system, which lacks the theoretical integrity of Dewey, but is more practical, is today used in nearly all large libraries in America. An example from one of the exceptions to this rule — the Purdue University Library — shows why. As taken from the Purdue catalog, the Dewey call number for Justin London's book *Modeling derivatives in C++* is 332.645701135262.

In the end, the card catalog showed the limits of alphabetical order. In very large files, even something as apparently simple as ABC… becomes quite complex. The invariant series it turns out is not necessarily invariant and often needs to be explained in some detail. How does punctuation file? What about spaces? Should "Mac" and "Mc" file together so that all of the "McDonalds" can be found in one place? Are numbers spelled out, and if not where do they file? What about abbreviations? The *ALA Filing Rules*, published in 1980 at the end of the card catalog era to explain such things was 50 pages

[11] McArthur, Tom, *Worlds of Reference*, page 121.

in length. The *Library of Congress Filing Rules* published the same year was more than twice as long.

As useful as they were, neither alphabetical order nor Dewey's classification system were sufficient to manage the knowledge that was being created by the middle of the twentieth century. The two means of bringing order to knowledge that had served reasonably well up to this time — thematic hierarchies and alphabetical order — were being strained even before the end of the print era. With the dawn of the Internet and the World Wide Web and the explosion of information they wrought new tools that would be needed to supplement what had worked in the past.

Hypertext

When we think about how to find things in cyberspace, today we inevitably think about Google, but before we get to Google, a short side trip to visit Ted Nelson, one of the great visionaries of the early computer era, is in order.

Theodor Holm Nelson was born in 1937, son of the Emmy Award winning director Ralph Nelson and the Academy Award winning actress Celeste Holm. He earned his undergraduate degree in philosophy from Swarthmore College and began his graduate studies in Sociology at Harvard in 1960. While at Harvard, Nelson took a course on the use of computers in the humanities. As a project for the course, he began work on a word processing system. Beginning a trend for which he has become legendary, the coding on the project was never finished, but as a result of this experience he began a life of imagining how the computer could transform the way knowledge was structured, created, and accessed. He coined the word "hypertext" to describe his non-sequential writing system and laid out his vision in a paper presented in 1965, at least two, and maybe three or four, decades before the technology was mature enough to realize it. But this did stop Nelson. Half prophet crying in the wilderness and half salesman he preached and sold his vision — which he called Xanadu — of how knowledge should be managed in the computer age. As Gary Wolf described it in a 1995 *Wired* article:

> Nelson's writing and presentations inspired some of the most visionary computer programmers, managers, and executives - including Autodesk Inc. founder John Walker - to pour millions of dollars and years of effort into the project. Xanadu was meant to be a universal library, a worldwide hypertext publishing tool, a system to resolve copyright disputes, and a meritocratic forum for discussion and debate. By putting all information within reach of all people, Xanadu was meant to eliminate scientific ignorance and cure political misunderstandings. And, on the very hackerish assumption that global catastrophes are caused by ignorance, stupidity, and communication failures, Xanadu was supposed to save the world.[12]

[12] Wolf, Gary, "The Curse of Xanadu," *Wired* 3.06 June 1995. http://www.wired.com/wired/archive/3.06/xanadu_pr.html (February 5, 2006).

Xanadu was never finished. Given Ted Nelson frenetic and disjointed personality, this is not a surprise, but Nelson's vision played a major role in shaping the work of Tim Berners-Lee and many others who have created the World Wide Web. Nelson is still the prophet crying in the wilderness. Of the Web, which he was largely responsible for inspiring, he has said, "The Web is the minimal concession to hypertext that a sequence-and-hierarchy chauvinist could possibly make."[13]

While Nelson is not satisfied, hypertext linking is central to the way the Web is organized and functions. In the print world citations were used to document connections, but finding the cited material was often difficult and always time consuming. Nelson's vision, that has been implemented, at least in an imperfect way, makes these links easy and immediate. Nothing needs to be looked up. You simply click on the link and go. This web of connections is so fundamental to the way we use the Web that it is assumed at an almost unconscious level. And, of course, it is now the name we often use for the Internet. How powerful is this linking structure? We have all heard of six degrees of separation, the notion that everyone on the planet is separated from everyone else by only six relationships. Albert-Laszlo Barabasi, a physicist from Notre Dame, reports in his fascinating book *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*, that the "diameter" of the Web — that is the number of clicks between any two documents — is nineteen.[14] At the time of his research in 1998 the Web had 800 million nodes. The power of a network of links is extraordinary. As Barabasi explains:

> What is important is that huge networks, with hundreds of millions or billions of nodes, collapse, displaying separation for shorter than the numbers of nodes they have. Our society, a network of six billion nodes, has a separation of six. The web, with close to a billion nodes, has a separation of nineteen… The natural question is: Why? How do networks achieve such a uniformly short path despite consisting of billions of nodes? The answer lies in the highly interconnected nature of networks.[15]

Hypertext linking allows documents on the Web, and the knowledge they contain, to become part of a network in which the power of network connections can help to guide us in finding and using the knowledge that is exploding on the Web.

Before moving on, I will give Nelson one last word. "Strange," he says, "nobody believes that God created computers. Therefore we are under no divine obligation to use them according to tradition. We are, in principle, free to start over. But most people do

---

[13] Nelson, Ted, "Ted Nelson's Computer Paradigm, Expressed as One-Liners," http://xanadu.com.au/ted/TN/WRITINGS/TCOMPARADIGM/tedCompOneLiners.html (February 5, 2006).

[14] Albert-Laszlo Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life,* London: Plum Book, 2003, pages 33-34.

[15] Albert-Laszlo Barabasi, *Linked,* page 34.

not dare think about it. I say it's high time."[16]

Yahoo and the Early Attempts to Organize the Web

The first finding tools for the Internet created hierarchies of various sorts. Gopher, created in 1991at the University of Minnesota — thus the name, did so in the days before the Web. Yahoo, one of the first successful, Web finding tools began in the same way. Developed by two Stanford University undergraduates David Filo and Jerry Yang, who wanted to keep track of their favorite websites, Yahoo started life as "Jerry's Guide to the World Wide Web." The company's official website explains the advent of its name, "The name Yahoo! is an acronym for 'Yet Another Hierarchical Officious Oracle,' but Filo and Yang insist they selected the name because they liked the general definition of a yahoo: "rude, unsophisticated, uncouth."[17]

The later source of the name is, of course, from Swift and *Gulliver's Travels,* but it is the former that belies the nature of the original tool. Srinija Srinivasan, Yahoo's fiftieth employee about whom the *New York Times* said, "she might be the world's richest librarian," was hired to lead its Web cataloging team in 1995.[18] Again this is indicative of the initial approach of the tool — it looked to past practice. While Yahoo has become a search engine, it began life as a hierarchical list. Dewey would have approved, and Nelson, we can be sure, hated it.

Lists and hierarchies as strategies for organizing the Web did not last long. These strategies were quickly overwhelmed by the volume of information on the Web. The problem with the Web is, of course, not finding the needle in the haystack, but rather trying to figure out which of the thousand of needles you have just discovered is the right one. Or, as it is often said, "retrieving information on the Web is like trying to drink from a fire hose." Early search engines used various strategies. Infoseek featured a very complex system of search modifiers, including Boolean operators. AltaVista devised a method to store every word of every page on the Internet in a fast, searchable index that ranked pages based on the occurrence of search terms. But neither approach, nor those of Lycos, Excite, or other early search engines were really successful at solving the key problem — how to find the most useful sites and rank them so that the best information appeared first in the list of retrieved sites.

Google

And then, along came Google.

---

[16] Nelson, Ted, "Somebody's Got to Disagree," http://ted.hyperland.com/notherview/ (February 6, 2006).
[17] "The History of Yahoo! - How It All Started...," Yahoo Media Relations, http://docs.yahoo.com/info/misc/history.html (February 12, 2006).
[18] Kaufman, Lesslie, "Whatever Happened to the Class of '93?; At Stanford, the Web Changes Everything Personal Business For Stanford's Class of '93, the Web Changes Everything The Programmer; A Physics Major Makes Good The E-Retailer Lots of Money But No Dates Personal Business The Nonconformist Looking Beyond The Internet Life The Web Editor; Feeling Fulfilled In Love of Work," *New York Times*, February 20, 2000. p. BU1.

Google began as a research project of two Stanford University graduate students Larry Page and Sergey Brin both born in 1973. Larry Page is the son of Carl Vincent Page, a professor of computer science at Michigan State University and Gloria Page, a computer-programming teacher. Sergey Brin was born in Moscow, his father was a mathematician and his mother an economist. In 1979, when Sergey was five, his family immigrated to the United States when his father received an appointment as a professor of mathematics at the University of Maryland.

The name Google is a play on the word "googol" which refers to the number represented by 1 followed by one hundred zeros. First incorporated as a private company in September 1998, on August 19, 2004 Google went public and at the end of the day had a market capitalization of $23 billion. Though the stock price has fallen some in recent days, the company is worth over $100 billion today. At its peak in early 2004, Google handled upwards of 85% of all search requests on the World Wide Web. It is estimated that today Google runs on 100,000 Linux servers and has 380 million unique users per month. The company's modest mission is to, "organize the world's information and make it universally accessible and useful."

Though the *Oxford English Dictionary* has yet to acknowledge the fact, "google" has entered the English language, and several others, as a verb meaning, "to do a search of the web."[19]

As described in the Wikipedia article on Google, Page and Brin:

> hypothesized that a search engine that analyzed the relationships between Web sites would produce better results than existing techniques… Convinced that the pages with the most links to them from other highly relevant Web pages must be the most relevant pages associated with the search, Page and Brin tested their thesis as part of their studies, and laid the foundation for their search engine."[20]

Google works so well not because it tries to analyze the contents of a particular page or work and place the page or work in some sort of scheme. Rather Google works because it analyzes the network and looks at which pages have the most links to them. When a search retrieves thousands of hits, the pages are ranked based on how well each page is known to, or is linked to, prominent pages. The items at the top of the ranked list are the most linked to and are as often as not, the most relevant. Google lists first those items that have the most and the best connections in the Web. It is as if the Web had voted and decided which pages where the most relevant.

Think of it this way, if you wanted to find a lawyer, you could go to the phone book and

---

[19] The *Oxford English Dictionary* (web edition) defines the verb "google" only as a cricket term, " *intr.* Of the ball: to have a 'googly' break and swerve. Of the bowler; to bowl a googly or googlies; also (*trans.*), to give a googly break to (a ball). Hence googler, a googly bowler," http://dictionary.oed.com (February 13, 2006).

[20] "Google – History," *Wikipedia: The Free Encyclopedia*, http://en.wikipedia.org/wiki/Google (February 12, 2006).

there would be a long alphabetical list. This is a start, but not really a very good one. If, on the other hand, the phone book worked like Google, it would know which lawyers in the list were in some way connected to widely known lawyers, like say our recent past President Henry Rider, and would arrange the lawyers not by the first letter of their last name, but rather by how well they were connected to Henry. This might not always be a perfect system, but it works much more often than not.

Conclusion: We are Clever Toolmakers

So what is the point of this excursion? What have we learned?

Before printing compilations of knowledge were arranged as simple hierarchical schemes that could easily be kept in one's head. With printing, knowledge expanded and new structures to arrange and retrieve it were required. Key among these was alphabetic order. The logic of theoretically beautiful organizing principles gave way to practicality.

When printing was industrialized in the nineteenth century again it outgrew the strategies used to organize and retrieve it. Dewey and others created professional library practice, and devised a series of innovations which where both practical and which possessed a theoretical basis. In the end the practical won out and the strategies that were developed were largely successful until the dawn of the computer age.

The Internet and Web brought a third revolutionary expansion of knowledge. As in the previous two cases, new tools were required to create and use this new container for knowledge. Hypertext and linking, and the monitoring of these links using Google's mathematical models are the most effective tools we have to date. It is probably too early to predict, but I suspect that whether Google's approach proves lasting or not, the tools that emerge in the future to make knowledge on the Web findable will use a similar strategy. These tools will watch and monitor the Web and use individual's interactions with it to provide pointers to the best information.

We are clever toolmakers. First we invent tools to increase our ability to create and communicate knowledge, then we have invent the tools required to manage all the new knowledge we have created. So far we have been largely successful. And while some of us might look at the Web in the same way many medieval scribes must have viewed the printing press, I don't think this is likely to change.